Aleck Greenham 20362627

Ash Tyndall 20915779

CITS3401 Data Exploration & Mining Project 1

Healthy Burgers Fast Food Chain

Abstract:

This document details the design of a data cube suitable for the *Healthy Burgers* franchise, a fast-food restaurant chain with stores in three countries, to aid management in its plans for improving sales and global expansion. System requirements were documented in *CITS3401 Data Exploration and Mining - Project 1* [1] ; reasonable assumptions were made and documented where the initial requirements were ambiguous or incomplete.

Dimensions:

* Stores
* Designs
* Facilities
* SalesPeriod
* PromotionalPeriod
* Years
* Combos
* Suppliers
* PriceRange
* CalorieRange

Methodology:

Palo for Excel [2]

References

[1] http://undergraduate.csse.uwa.edu.au/units/CITS3401/labs/proj1-2013.html

Overview:

- Burger franchise

- interested in improving sales and popularity

- each franchise has own OLTP database

- state, country and worldwide DB exist

- **countries**: Australia, New Zealand, Singapore - 3 **store**s each

- **Sales period**s: breakfast, lunch, dinner

- 6 different **promotion**s a year

- 6 different **combo meals**, not stocked in each store

- combo meals divided into **price categories**

- each sale period has 2 combo meals

- publish **calorie value** of each burger (high calorie burgers reported as tasting better)

- 2 different **supplier**s

- each supplier does **ingredient**s for 3 combo meals (each combo meal only has 1 supplier)

- 3 different **interior designs**

- 3 different **facility types**: drive-through, dine-in, both

- historical data between 2008 and 2012

|  |  |
| --- | --- |
| **Deliverable** | **Details** |
| Schema & Data Cube Details Document |  |
| 5 Screenshots of Example Scenarios | - should be “interesting or important” |
| Data for populating schema | - Interested only in sales figures in dollars |
| Data Cube Demonstration |  |

**TODO**

1) ~~Get up-to-date with lecture notes~~

2) ~~Convert requirements into schema~~

3) I~~nvestigate & install Palo~~

4) ~~Decide how to generate data~~

5) ~~Decide on valuable analysis to take screenshots of~~

6) Write a document containing schema and interesting analysis

7) Create a list of anticipated questions

Schema Document

**Requirements Summary**

- interested in improving sales and popularity

|  |  |  |  |
| --- | --- | --- | --- |
| **Object** | **Properties** | **Relationship** | **Limitation** |
| Country |  |  | Australia, New Zealand, Singapore |
| Store | - Interior Design  - Facility Type | - Country | - 3 stores in each country  - 3 different interior designs  - facilities can be dine in, drive through or both |
| Combo Meal | - Price Category  - Calorie Value | - Supplier | - 6 different combo meals  - not all stores have each combo meal  - divided into 3 different price categories |
| Supplier | - Name |  | - there are 2 suppliers  - does ingredients for each 3 combo meals |
| Sale | - Date  - product | - Sales Period | Date between 2008 and 2012  Sales Period must be Breakfast, Lunch, Dinner |
| Sales Period |  |  | Has 2 combo meals |
| ~~Promotion~~ |  |  | ~~6 each year~~ |
|  |  |  |  |
|  |  |  |  |

Assumption 1: The type of ingredients are not important, only their supplier - i.e. the supplier is a proxy for how desirable the ingredients they supply are

~~Assumption 2: Won’t normalise sale periods because Datawarehouse does not need to update type of sales periods (will stick with breakfast, lunch, dinner). If you were to add an additional sales period, this would not change existing data. Data parsing would sort out if the same sales periods had a different name (dinner == teatime)~~

~~Assumption 3: Price categories should not be normalised, as it is easier for data analysis if they are stored with the combo meal~~

~~Assumption 4: Sales period should be stored on combo meal so only have to be stored once~~

Assumption 5: high calorie burgers reported as tasting better is irrelevant information

Assumption 6: Sale periods are identified by their names and not by time ranges.

Assumption 7: There are six combo meals globally

Assumption 8: For there to be 3 sales periods, each with 2 combo meals, and there only being a total of 6 combo meals to choose from. If it’s true that not all stores sell all the combo meals, then some combo meals must be shared between sales periods.

Assumption 9: All non-combo products sold are not of interest (because they have no relation with any of the other data fields)

Assumption 10: The same suppliers supply the ingredients for same combo meals, globally.

~~Assumption 11: Promotion refers to combo meals.~~

Assumption 12: The price of combo meals is set at the store level

Assumption 13: A promotional period has it’s own 6 combo meals (thus the combo meals change every 2 months)

**Attempt 1: Star Schema**

**Design Justification:**  The assignment sheet states that the burger chain is only interested in sales, which removes the need to have multiple fact tables (a constellation schema).

**Basic principles**: A Data Warehouse should consist of:

1) A Fact Table that contains keys to all the dimensional tables **and** measure fields (eg. units sold)

2) A series of dimensional tables that group different levels of abstraction of the same thing

**Working From:**

|  |  |
| --- | --- |
| **Table** | **Fields** |
| fact | restaurant\_key  combod\_meal\_key  time\_key  location\_key  **Measures:**  units\_sold  dollars\_sold  avg\_sales |
| restaurant | restaurant\_key  interior\_design  facilities |
| combo\_meal | combo\_meal\_key  price\_category  calorie\_value  supplier |
| time | time\_key  year  sales\_period  promotional\_period |
| location | loaction\_key  country |

**Unknowns:**

- Where are the sales stored?

- Are the measures calculated ahead of time, or are they dynamic (dependent on OLAP operations such as drill down, pivot, etc)?

**Instructions:**

“... submit screenshots of **five** *example scenarios* that you think are **interesting** and **important**.”

**Guiding Info:**

“Healthy Burgers plans to **expand the business further around the world**...”

“... for **improving sales and popularity** of its products.”

“Healthy Burgers wants to evaluate a prototype of its data warehouse for three countries ...”

“... interested in a **thorough analysis of historical data** between 2008-2012”

“We are **interested in** only one measure, **sales figures in dollars**.”

**Unknowns:**

Definition of “example scenarios”

**Overarching Metric:**

Sales figures (dollars)

**Breakdown:**

|  |  |  |
| --- | --- | --- |
| **Objective** | **Relevant, Possible Correlation Between Sales Figures and ...** | **Scenario** |
| Global Expansion | - Country  - Restaurant Facilities |  |
| Improving Sales & Popularity of Product | - Calories Content  - Price  - Supplier  - Sales period |  |
| Analysis of sales evolution between 2008-2012 |  |  |
|  |  |  |

**Behaviour We Expect To See In The Data**

1) Restaurant facilities should determine total number of sales made across all products. For example, a drive through facility may bring in more customers, which will increase sales of all products equally (assuming people do not purchase meals easier to consume in their cars).

2) The relative sales of an individual product should depend on price, calorie content, supplier (quality of ingredients) and sales periods (whether it is suitable meal for breakfast, lunch or dinner)

3) Sales over a two year period of a single restaurant (assume all restaurants are older than two years) will depend on factors not measured such as advertising campaigns, community awareness of the outlet, seasons, etc.

4) Global expansion should depend entirely on factors not measured, such as market research into the countries the company is planning on moving into. All of this data is useless for that purpose and their objective should be dismissed or ignored. The data gathered is only useful for improvements of existing restaurants, assuming a reasonably homogenous consumer base across all three countries, and what works in one country or restaurant can be similarly applied in another country or restaurant with like effects.

**Thinking Out Loud:**

A business looking to expand in a global market is interested in what works, for the purpose of replicating the success elsewhere. Although what determines a burger chain restaurant’s success is complicated and realistically outside the scope of just the variables reflected in this assignment, data of some appreciable utility would be:

As you cannot reproduce a country or a particular restaurant (as an atomic unit), these have been removed from the possible total sales column.

|  |  |  |
| --- | --- | --- |
| **Variable of Interest** | **Correlated Factor** | **Description** |
| Total Sales | - per restaurant facility  - per restaurant layout  - per price range | - Which restaurant facilities had the greatest sales?  - Which restaurant layout had the greatest sales?  - Which price range of products made the company the most money? |
| Product Sales | - per sales period  - per restaurant facility | - Which sales period sold the most products?  - Is there a link between restaurant facilities and the number of a certain product so? |